Regular Article

Decentralized Online Social Network Using Peer-to-Peer Technology

Manh Ha Tran, Van Sinh Nguyen, Synh Viet Uyen Ha

School of Computer Science and Engineering, International University, Ho Chi Minh City, Vietnam

Correspondence: Manh Ha Tran, tmha@hcmiu.edu.vn

Abstract– **This paper presents a social network with a peer-to-peer architecture that facilitates social computing services in distributed environments. This social network aims to provide users the capability of managing the dissemination of user data, searching user data on the data silos of the network, and consolidating user data from various social networks. The social network employs a super peer peer-to-peer architecture that contains peers and super peers. Users use peers to participate the network and services. Peers with sufficient storage, bandwidth and processing power become super peers that support peers for complex operations such as user authentication or group communication. We have extended the Gnutella protocol to provide the authentication and posting services on the social network. The design of these services copes with the distributed setting of the social network. The evaluation of the prototyping social network has performed on a number of laboratory workstations to investigate its scalability, reliability and performance.**

Keywords– **social network, peer-to-peer network, decentralized online social network, posting service, authentication service.**

1 INTRODUCTION

Social networks today connect millions online users and contribute a significant amount of digital content to the Internet. Social and social media networks such as Facebook [1], MySpace [2], Twitter [3], Flickr [4], Youtube [5] continue to grow and attract more users in the forthcoming years. The fast development of social networks not only brings benefits to users but also poses several problems [6]. First, users cannot manage the dissemination of their personal data on social networks. Second, users cannot search data on the data repository of social networks. Finally, users cannot consolidate their personal data from various social networks. One of the main reasons that causes these problems is the dependence of the client-server architecture with centralized and proprietary servers.

We propose a social network that uses a super peer peer-to-peer (P2P) architecture for facilitating social computing services in distributed environment. The P2P based social network inherits several good characteristics of the P2P architecture including scalability in architecture, reliability in content distribution and autonomy in administration. This social network can exploit the data sharing and searching capability of the P2P architecture to provide social computing services. We then introduce two authentication and posting services on this social network. The authentication service reduces the dependence of centralized servers, and the posting service improves the group communication of users. The study in this paper thus focuses on:

- Proposing a social network based on a P2P architecture supporting social computing services on distributed environment;
- Extending the P2P architecture and protocol to facilitating social network's functions including user group management and communication;
- Implementing two authentication and posting services for the proposed social network using the Gnutella protocol [7];
- Evaluating the scalability, reliability and performance of the prototyping implementation using a number of laboratory workstations.

The rest of the paper is structured as follows. The next section includes the background of P2P networks, a brief overview of social networks, and description of some existing approaches of decentralized online social networks. Section 3 proposes a social network architecture using a super peer P2P network. This section describes two authentication and posting services that have been designed to operate on this social network. It also presents several implementation details of the prototyping social network using an open source package of the Gnutella protocol. Several experiments in Section 4 report the scalability, reliability and performance of the prototyping implementation before the paper is concluded in Section 5.

2 RELATED WORK

A P2P network is a collection of networked computers and mobile devices referred as peers. Peers acting as both client and server share computing resources including file, storage, bandwidth and processor power through consuming and provisioning services, respectively. The P2P network is established by a special communication process that allows peers to join and leave
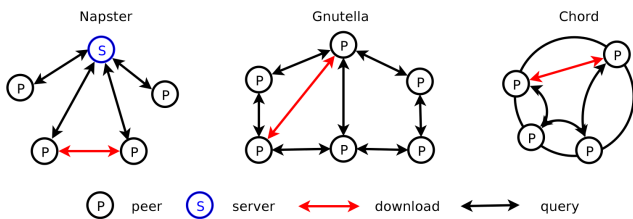
Figure 1. Searching and sharing resources on Napster, Gnutella and Chord.

dynamically with some degree of self-administration, fault-tolerance and scalability. Considering the mechanisms of searching and sharing resources, P2P networks can be divided into three categories: centralized, structured and unstructured networks.

A centralized P2P network uses a centralized server to index resources from peers. A peer searches the resource indexes and corresponding peers from the server and directly downloads the resources from the corresponding peers. The typical network includes Napster [8]. A structured P2P network uses Distributed Hash Table (DHT) to generate uniquely consistent identifiers for peers and resources such that the peers hold the resource indexes if their identifiers are in the same identifier space. Peers forward queries to other peers closer to the resource indexes in the identifier space to figure out the corresponding peers. The typical structured networks include CAN [9], Chord [10], and Kademlia [11]. An unstructured P2P network maintains resource indexes on peers. Peers search resources by flooding queries to other peers on the network, and the peers holding the resources send back their information for the download process. The typical unstructured P2P systems include Gnutella [7], Freenet [12], and BitTorrent [13]. In addition, a super peer P2P network is a hybrid network that combines the characteristics of the P2P network with the client-server network to address the problem of heterogeneous peers, i.e., peers possess various capability of storage, bandwidth and processing power. The study of Yang *et al.* [14] has presented guidelines for designing the super peer network to take advantage of peer capabilities. The super peer network comprises many clusters connected to each other to form either structured or unstructured P2P networks, in which each cluster contains a super peer and a set of clients. The clients submit queries to, and also obtain queryhits from, their super peer while the super peers forward the queries and receive the queryhits on the super peer network.

A study of Boyd *et al.* [15] has referred to social networks as networks that support online social activities including making friends online, publishing user data online through posting, commenting, chatting, etc. Social networks have offered a number of social services ranging from social news, wikis to multimedia sharing and social shopping. An online statistics specifies the dominance of the posting and chatting services among various services. Typical social networks that have attracted a huge number of users include Facebook [1], MySpace [2], Twitter [3], Flickr [4], Youtube [5]. These networks share some common characteristics: using the client-server architecture and web technology, thus suffering from the problems specified above. Popular social network activities on Facebook and Twitter can be classified into three categories:

- Message exchange allows users to send and receive short text update using chat, blog, forum, etc.
- Resource sharing allows users to upload and download multimedia data such as photos, video, ebooks, etc.
- Browsing event allows users to search data, make friends, play online social games, etc.

Decentralized online social networks provide various social computing services on distributed environment. These services improve the limitations of the centralized servers by using the decentralized servers. P2P technology has successfully been applied to some classes of applications including content delivery, file sharing, multimedia streaming, search and lookup. Note that these applications are beneficial to social networks. P2P systems have proved to solve the limitations of the client-server architecture. Several studies have applied P2P technology to building decentralized social networks. Safebook [16] adopts a decentralized architecture relying on cooperation among users to deal with the user privacy and provider application. It also builds trust relationship between users for using online applications. LifeSocial [17] tackles the problem of high administration cost on multimedia online communities that heavily depend on the centralized systems. It uses a P2P framework to construct a network architecture and extend functions to include social networking services, such as user profiles, friend lists, groups, photo albums, etc. This framework has later been applied for building secure online social networks. PeerSoN [18] resolves the problem of user privacy and Internet connectivity on social networks. It replaces the centralized authority of social networks and provides direct data communication between network nodes by using a P2P framework with encryption. It also uses distributed storage to foster local services. Some of the studies have provided the prototyping systems and API interfaces for implementing services. The study of Tran *et al.* [19] has applied mobile computing for social network services based on the increasing demand of using mobile devices for Internet services. They have proposed a mobile P2P approach with network formation and search service that allow users to participate in data sharing and searching activities. The approach uses the Gnutella protocol and software bug report datasets for evaluation. However, these studies only deal with some parts of the decentralized social network, either architecture or service issues. LifeSocial includes a complete framework of the decentralized social network with unclear implementation details. The study in this paper contains the proposal and implementation of a decentralized social network featured by the Gnutella-based architecture with the authentication and posting services.
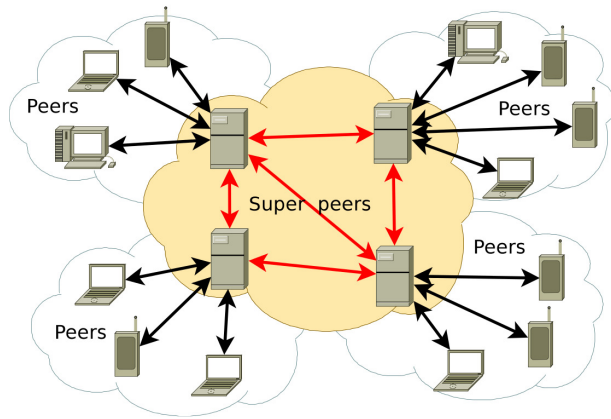
Figure 2. An architecture of P2P based social network.

3 P2P Based Social Network

A P2P based social network employs a super peer P2P architecture to achieve remarkable features in distributed environment: scalability in architecture, reliability in content distribution and autonomy in administration. The architecture also deals with the problem of peer heterogeneity by using peers with sufficient storage, bandwidth and processing power to support other peers for complex operations. There are two kinds of peers in this architecture: super peer and peer as shown in Figure 2.

Super peers possess not only sufficient system resources and stable connectivity but also high uptime without being behind a firewall. Each peer with normal system resources connects to some super peers, while super peers connect to other super peers. Super peers act as proxies for peers and get involved in message routing mechanisms, i.e., messages are stored and forwarded among super peers. When receiving a message, a super peer choose to forward the message to other super peers or deliver the message to a peer that matches the message's condition. Peers can become super peers if they have sufficient capability. We have proposed two fundamental social services: authentication and posting for the P2P based social network. The authentication and posting services feature the capability of reducing the dependence of the centralized servers and fostering group communication in the social network.

3.1 Authentication Service

Users using peers can join the social network without authentication as the way users join the P2P network. However, users need authentication to use services on the network, e.g., posting service. This network supports several published registration servers for user registration. After successful registration, users receive a list of super peers for connecting to the network, and the registration server also informs the same list of super peers about the registered users. For the next joining time, users gain access to the network by authenticating with some super peers. If super peers are offline, users can authenticate with the registration

---

Algorithm 1 User's Authentication Process

1: **void** autenticateUser (String userAccount, String listSuperPeers)
2:   **if** user logics for the first time
3:     send userAccount to registration server
4:     receive listSuperPeers from registration server
5:   **end if**
6:   **while** number of super peer connections is not maximum
7:     **if** a super peer is alive in listSuperPeers
8:       send userAccount to the super peer for authentication
9:     **else**
10:       update listSuperPeers from registration server
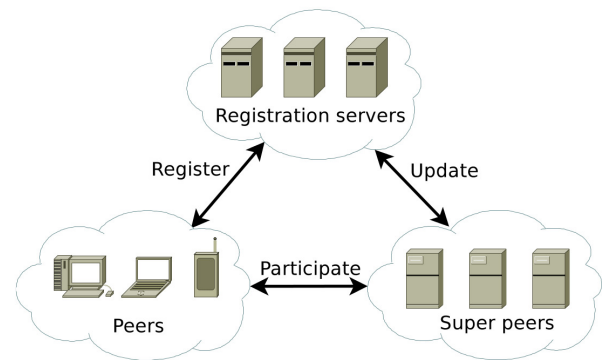11:   **end if**
12:   **end while**

Figure 3. Social network joining process.

servers and obtain an updated list of super peers, as shown in Algorithm 1.

The registration server contains a database that stores super peers and registered users, while super peers also contain a database that stores registered users synchronized by the registration server. Users can discover further super peers through social activities.

This authentication approach has been applied to some P2P applications, such as Skype [20]. Figure 3 describes the social network joining process that contains registration, update and participation. This authentication service allows users to join the network by authenticating with some super peers, thus reducing the dependence of the centralized login servers.

3.2 Posting Service

After joining the network, users using peers can participate the posting service that allows users to post messages to, and request messages from, individual users and user groups. Messages contain user status, user profile and discussion update. We consider updating user status and user profile as checking alive peers, where peers use short messages to discover P2P networks, e.g., a peer in the Gnutella network regularly send the ping short message to other peers and receive the pong message that contains the required information of the peers. We support group communication for updating discussion. Note that communicating with individual users is a special case of communicating with user groups. To implement group communication,
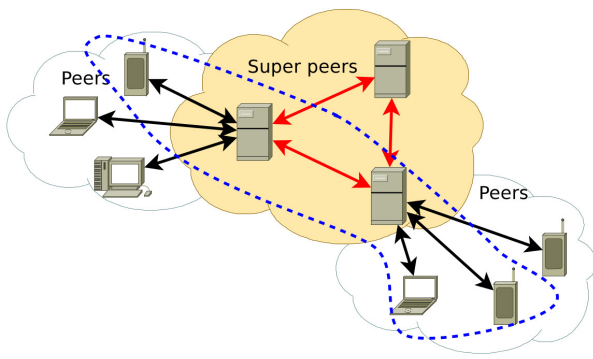
Figure 4.   Group message posting process.

Algorithm 2 User's Status Update Process

1: **void** postStatus (String userStatus, String listSuperPeers, String listFriends)
2:    save userStatus to user's profile in database
3:    creat a message to update super peers and friends
4:    **if** connection with a super peer exists
5:      send this message to the social network
6:      **if** user's peer is super peer
7:        save this message to the newsfeed file of user's friends
8:      **else**
9:        Put this message in a queue to resend later
10:    **end if**

---

the information of user groups is defined on user profiles, and users can select to send messages to the whole groups, a set of users or one user. Users then pass the messages and list of peers to the corresponding super peers that manage to deliver the messages to other peers. There is a scalability problem that users might not find out a certain peer on a large network of super peers. Group communication is a feature of this social network.

Group communication has been applied to some P2P applications [21]. Figure 4 depicts the group message posting process that allows a mobile device from the upper cluster to post a message to another mobile device and laptop from the lower cluster through 2 super peers. In this example, the mobile device decides on a group of users and the two super peers determine the routing mechanism to deliver the message to the other mobile device and laptop. This posting service allows users to keep their personal data on peers and super peers, thus improving data privacy and search capability on the network.

Algorithm 2 presents a process to update user's status. When a user writes a status on the user's peer, the status is stored in the user's profile, and a message is posted to the social network. This message contains the status, a list of super peers that maintain the status, and a list of friends who receive the status. If the user's peer connects to super peers, the message is sent to the social network through super peers, otherwise this message is stored in a queue to resend later. If the user's peer is also a super peer, this message is saved directly in the newsfeed file of the user's friends.
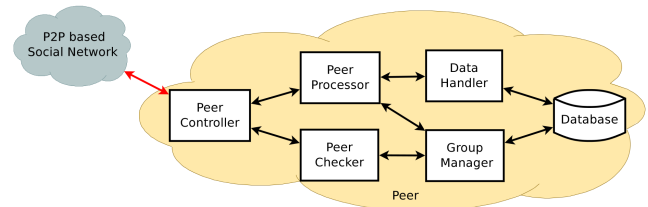

Figure 5.   Peer components and communication in the P2P based social network.

3.3 Implementation

A peer in the P2P based social network contains several components as shown in Figure 5. Peer controller is responsible for peer communication. This component allows peers to exchange various types of messages with other peers, e.g., messages for checking peer identities, messages for checking peers alive, messages for joining peer groups, messages for posting peer discussions, etc. Peer checker takes a role of registration servers to authenticate peers. This component can request a certain user registration information from the registration servers or super peers. Peer processor keeps track on peer activities related to the posting service, e.g., managing peer information including status and profile, processing posting messages, communicating peer groups, etc. This essential component closely works with the data handler and group manager components. Group manager is responsible for group formation and management. This component maintains the stability of various peer groups and facilities the data exchange of peers. Group communication is a main feature of the posting service. Data handler updates and stores messages on peers during peer activities. There are several kinds of messages related to the posting service. If a peer cannot deliver messages to other peers, it stores and forwards the messages later.

Both peers and super peers contain these components. However, the functions of the components can be different on peers and super peers. The peer checker component of the super peer is associated with user registration information stored on the database, while this component of the peer only contains itself registration information and a list of super peers. The data handler and group manager components of the peer is much simpler than those of the super peer. The function of storing and forwarding messages later is only suitable for the super peer.

The P2P based social network uses super peers to perform the authentication and posting services. The Gnutella protocol [7] is therefore extended to enable these services on super peers. The ping and pong messages have been changed to include the update information of peers such as user status and profile. The query and queryhit messages have been changed to search for peers and data. We have used several new messages for the posting service: post, like and comment messages to support creating and sharing activities on peers and groups, and responding to these activities; newsfeed request and response mes-

sages to support retrieving activities from other peers and groups, and responding to these activities; profile request and response messages to support obtaining profile information from other peers and groups, and responding to these activities; group and join messages to support managing peer groups and group communication. In addition, several messages that have been developed for the authentication service and peer data synchronization are unrelated to this protocol. We have applied these changes to the Java open source implementation of the Gnutella protocol. Each peer also contains a MySQL [22] database to store user data, peer data and messages.

4 Evaluation

Among several evaluation requirements of the P2P based social network, we choose to focus on scalability, reliability and performance rather than security. First, we have applied the Skype authentication service to this social network with a similar super peer P2P architecture, security issues can be reduced considerably. Second, the social network still requires a lot of improvement and optimization, thus performance issues are less important than scalability and reliability considering the posting service. For scalability, we measure the capability of the network to extend to a large number of peers and super peers in the dynamic environment where workstations can join and leave the network arbitrarily. For reliability, we measure the influence of peer and super peer failure on the operation of the authentication and posting services. For performance, we compare time consumption for different activities, such as network joining, newsfeed update and user registration.

We have established a P2P based social network based on 50 HP Pro workstations with Intel(tm) Core i3 Processor 3.30GHz, 2GB RAM, 512GB HDD running at the networking and database laboratories of our university. We have used XEN virtualization software [23] to create a number of limited virtual machines for peers and super peers. Each peer can connect to 3 super peers and each super peers can connect to 10 other super peers and accept 20 peers at maximum. The network is simulated with some parameters including packet lost, packet delay and packet reorder in order to create a similar Internet environment. Figure 6 on the left side reports memory usage for storing messages on peers and super peers. While peers only store their messages and group messages, super peers take a role of intermediate storage for storing and delivering peer messages. Since the average message size is small, peers need trivial amount of memory, and super peers need 200MB approximately to store 10 thousand messages. Ordinary configuration of workstations and mobile devices can easily adapt with this memory requirement.

Figure 6 on the right side presents traffic generation for processing messages on peers, super peers and Gnutella peers. The appearance of Gnutella peers is used for comparison because the Gnutella protocol uses the message flooding mechanism that causes large amount of traffic on the network. When receiving a message, both Gnutella peers and super peers react similarly by flooding the message to all the neighboring peers, and peers only forward the message to some super peers. Peers generate less traffic to the network comparing to Gnutella peers and super peers, i.e., 18 MB comparing to 44 MB over 10 thousand messages. Gnutella peers generate the same amount of traffic as super peers. This experiment also considers the uncertainty of the Gnutella and super peer networks. Heterogeneous peers on the Gnutella network are less stable than super peers that possess sufficient capability of performing complex operations.

Figure 7 on the left side depicts the capability of super peers to handle peers in dynamic environment where peers can join and leave the network arbitrarily. Churn rate has been widely used in the P2P network community to present the number of peers moving out and in the network over a certain period of time. This experiment measures number of peers and super peers with various churn rates. We have used 10% to 40% for peers and 2% to 4% for super peers. The stability of peers reduces as churn rates and number of super peers increase, i.e., the line of peer 10% is stable, while other lines is unstable. The line of peer 20% performing similar to the line of peer 30% indicates super peers have some impact on the stability of peers even the churn rate of the super peer network is low.

Figure 7 on the right side reports the failure probability of super peers in two situations: failure with reconnection and failure without reconnection. Super peers in the former situation can go online again after disconnection, e.g., a super peer can disconnect to some super peers and connect to other super peers for improving performance, while super peers in the later situation may encounter software or network problems, causing data unavailability on the network. Since high uptime is one of criteria to select super peers, the possibility of super peers to go offline is thus low except for unexpected reasons, i.e., 10% to 15% of super peers can fail without reconnection in this experiment. Note that this rate is similar to churn rate. However, the possibility of super peers to go offline and online is high, i.e., 40% to 50% of super peers can fail and reconnect during the experiment period.

Figure 8 compares time consumption for three activities. First, a peer joining the social network requires a list of super peers updated by the registration server to open connections to super peers. Second, a peer gets newsfeed update from other peers and peer groups stored on super peers. Third, a peer registers with the registration server that provides the peer's information update to a list of super peers and also sends this list to the peer. There are two scenarios in this experiment: a peer connects to 2 super peers with 10% churn rate and a peer connects to 3 super peers with 20% churn rate. An increasing number of super peer connections can alleviate failure impact on the social network.

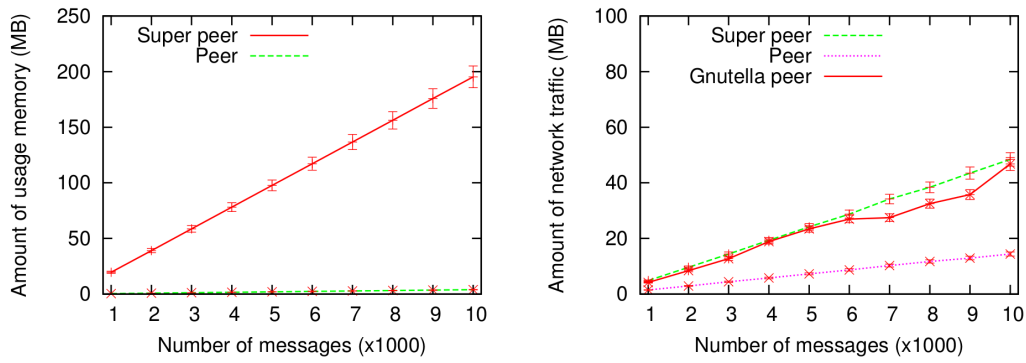Figure 8 on the left side presents the first scenario

Figure 6.   Memory usage for message storage (left). Traffic generation for message processing (right).
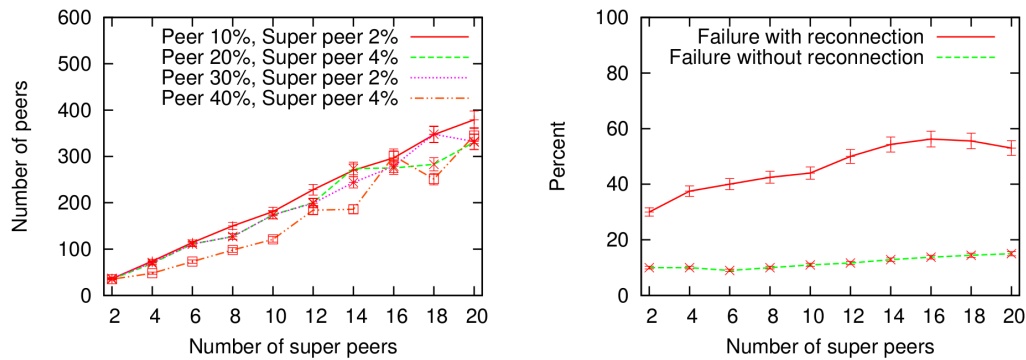

Figure 7.   Number of peers and super peers over various churn rates (left). Failure probability of super peers (right).
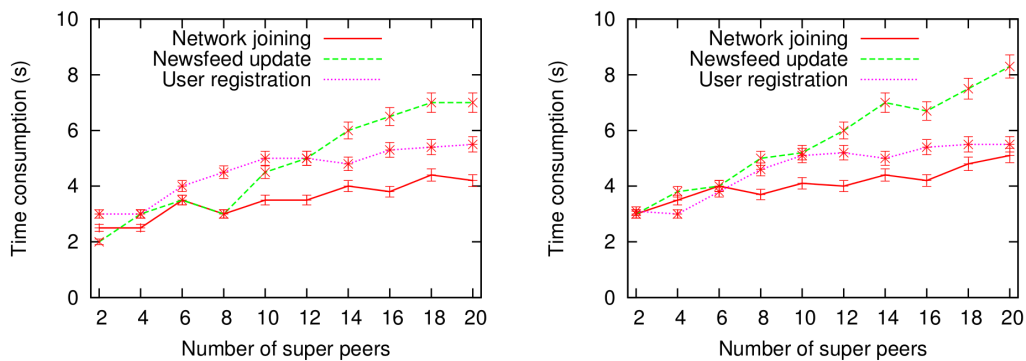

Figure 8.   Comparison of time consumption for different activities with 10% churn rate and 2 super peers connection (left). Comparison of time consumption for different activities with 20% churn rate and 3 super peers connection (right).

considering 10% churn rate and 2 super peer connections. Time consumption for newsfeed update linearly increases while time consumption for network joining and user registration slowly increase as a number of super peers increase. It takes time for super peers to communicate multiple super peers and get newsfeed update for peers. Increasing super peers in the social network causes trivial impact on network joining because peers only choose 2 or 3 super peers from a list of super peers for connection. Time consumption for user registration also depends on the load balance of super peers when providing the peer's information update to a list of super peers. Time consumption for network joining and user registration can be affected by failure rate in the network.

Figure 8 on the right side presents the second scenario considering 20% churn rate and 3 super peer connections. Newsfeed update also consumes a lot of time compared to network joining and user registration as a number of super peers increase. Since churn rate increases, it takes more time for super peers to obtain newsfeed update due to intermittent connections. Connecting to 3 super peers also causes impact on network joining because peers must connect to further super peers, and high churn rate can cause connection failure. Time consumption for user registration is rather stable for both scenarios because the registration server usually provides extra super peers for registered peers. It only consumes a lot of time when several super peers go offline at the same time, but this situation is rare.

5 Conclusions

We have proposed a P2P based social network that fosters social computing services on distributed environment, particularly authentication and posting services. This social network allows users not only to manage the dissemination of personal data, to search user data on the data silos of the social network, but also to consolidate user data into multiple social networks. This social network possesses a P2P architecture that contains peers and super peers. Users register and participate the network as peers, where peers with sufficient capability of storage, bandwidth and processing power can become super peers that are responsible for complex operations. For the authentication service, peers can authenticate with super peers instead of centralized servers before joining the network. For the posting service, super peers can store and forward messages to groups of peers.

We have used the Gnutella protocol to implement the prototyping social network. The experiments evaluate the scalability, reliability and performance of the social network using several metrics: memory usage, traffic generation, peers and super peers availability and failure, time consumption for different activities. The experimental results reveal that super peers with normal configurations can provide sufficient memory for a number of peers and generate as the same network traffic as Gnutella peers, while peers use trivial memory for storing messages and generate low amount of network traffic for processing messages. Super peers possess high uptime and reconnect the network after failure, while the high churn rate of peers cause the instability of the network. Moreover, time consumption for different activities is stable and reasonable for the current configuration of the social network. The experiments have been performed on a number of laboratory workstations. Future work considers the improvement of the social network's architecture for better performance and security, and the possibility of extending the social network to other social services, such as audio and video conferencing services, photo sharing services.

References

[1] Facebook Social Network, http://www.facebook.com/. Last access in Oct. 2014.

[2] Myspace Social Network, http://www.myspace.com/. Last access in Oct. 2014.

[3] Twitter Social Network, http://www.twitter.com/. Last access in Oct. 2014.

[4] Flickr Social Media, http://www.flickr.com/. Last access in Oct. 2014.

[5] Youtube Social Media, http://www.youtube.com/. Last access in Oct. 2014.

[6] C.-m. A. Yeung, I. Liccardi, K. Lu, O. Seneviratne, and T. Berners-Lee, "Decentralization: The future of online social networking," in *W3C Workshop on the Future of Social Networking Position Papers*, vol. 2, 2009, pp. 2–7.

[7] Gnutella Protocol Specification 0.4, http://rfc-gnutella.source forge.net/developer/stable/index.html, last access in Mar. 2014.

[8] B. Carlsson and R. Gustavsson, "The rise and fall of napster - an evolutionary approach," in *Proceedings of the 6th International Computer Science Conference on Active Media Technology (AMT'01)*, ser. AMT '01.   London, UK: Springer-Verlag, 2001, pp. 347–354.

[9] S. Ratnasamy, P. Francis, M. Handley, R. Karp, and S. Schenker, "A Scalable Content Addressable Network," in *Proc. Conference on Applications, Technologies, Architectures, and Protocols for Computer Communications (SIGCOMM '01)*.   New York, NY, USA: ACM Press, 2001, pp. 161–172.

[10] I. Stoica, R. Morris, D. Karger, M. F. Kaashoek, and H. Balakrishnan, "Chord: A Scalable Peer-to-Peer Lookup Service for Internet Applications," in *Proc. Conference on Applications, Technologies, Architectures, and Protocols for Computer Communications (SIGCOMM '01)*. New York, NY, USA: ACM Press, 2001, pp. 149–160.

[11] P. Maymounkov and D. Mazières, "Kademlia: A Peer-to-Peer Information System Based on the XOR Metric," in *Proc. 1st International Workshop on Peer-to-Peer Systems (IPTPS '01)*.   London, UK: Springer-Verlag, 2002, pp. 53–65.

[12] I. Clarke, O. Sandberg, B. Wiley, and T. W. Hong, "Freenet: A Distributed Anonymous Information Storage and Retrieval System," in *Proc. International Workshop on Design Issues in Anonymity and Unobservability*. Heidelberg, Germany: Springer-Verlag, 2000, pp. 46–66.

[13] B. Cohen, "Incentives Build Robustness in Bittorrent," in *Proc. 1st Workshop on Economics of Peer-to-Peer Systems*, 2003.

[14] B. Yang and H. Garcia-Molina, "Designing a super-peer network," in *19th International Conference on Data Engineering, 2003. Proceedings*.   IEEE, 2003, pp. 49–60.

[15] D. M. Boyd and N. B. Ellison, "Social network sites: Definition, history, and scholarship," *Journal of Computer-Mediated Communication*, vol. 13, no. 1, pp. 210–230, 2007.

[16] L. A. Cutillo, R. Molva, and T. Strufe, "Safebook: A privacy-preserving online social network leveraging on real-life trust," *IEEE Communications Magazine*, vol. 47, no. 12, pp. 94–101, Dec. 2009.

[17] K. Graffi, C. Gross, D. Stingl, D. Hartung, A. Kovacevic, and R. Steinmetz, "LifeSocial.KOM: A Secure and P2P-based Solution for Online Social Networks," in *Proc. IEEE Consumer Communications and Networking Conference (CCNC)*.   IEEE, 2011.

[18] S. B. D. Schiöberg, L. H. Vu, and A. Datta, "PeerSoN: P2P social networking: early experiences and insights," in *Proc. 2nd ACM EuroSys Workshop on Social Network Systems (SNS '09)*.   New York, NY, USA: ACM, 2009, pp. 46–52.

[19] H. M. Tran, K. V. Huynh, K. D. Vo, and S. T. Le, "Mobile peer-to-peer approach for social computing services in distributed environment," in *Proceedings of the Fourth Symposium on Information and Communication Technology*. New York, USA: ACM, 2013, pp. 227–233.

[20] Skype VoIP, http://www.skype.com/. Last access in Oct. 2013.

[21] D. Tsoumakos and N. Roussopoulos, "Agno: An adaptive group communication scheme for unstructured p2p networks," in *Proc. 11th International Euro-Par Conference*. Springer, 2005, pp. 1183–1193.

[22] MySQL Database Software, http://www.mysql.com/. Last access in Mar. 2013.

[23] P. Barham, B. Dragovic, K. Fraser, S. Hand, T. Harris, A. Ho, R. Neugebauer, I. Pratt, and A. Warfield, "Xen and the art of virtualization," *ACM SIGOPS Operating Systems Review*, vol. 37, no. 5, pp. 164–177, Oct. 2003.

Ha Manh Tran is a lecturer of the School of Computer Science and Engineering at International University - Vietnam National University. He received his master degree of computer science in 2004 from the University of Birmingham, United Kingdom and his doctoral degree of computer science in 2009 from Jacobs University Bremen, Germany. His research interests include distributed computing, big data analytics, information retrieval and network management.

Sinh Van Nguyen is a lecturer of the School of Computer Science and Engineering at International University-Vietnam National University. He received his master degree of computer science in 2008 from Asian Institute of Technology, Thailand and his doctoral degree of computer science in 2013 from the School of Doctorate in Mathematics and Information Technology, Aix-Marseille University, France. His research interests include big data analytics, web technology, source code security, images processing.

Synh Viet Uyen HA is a lecturer of the School of Computer Science and Engineering at International University - Vietnam National University. He received his master degree in 1999 from University of Natural Science, Vietnam and his doctoral degree in 2010 from School of Information and Communication Engineering, Sungkyunkwan University, Korea. His research interests include machine learning, computer vision, image processing.